

Claims

What is claimed is.

1. A text information generating apparatus, comprising:
 - attribute input section operatively connected to receive at least one artificial attribute associated with a paragraph;
 - discourse structure attribute generating section operatively connected to generate a discourse structure attribute related to a discourse structure that is associated with said paragraph and a paragraph length ratio attribute related to a ratio of a number of characters in said paragraph to the number of characters of a matching pattern matched with said paragraph;
 - combination attribute generating section operatively connected to generate a combination attribute based on at least two of the artificial attribute, the discourse structure attribute, and the paragraph length ratio attribute;
 - text input interface operatively connected to receive text;
 - importance degree estimating section operatively connected to estimate an importance degree indicating an enhancement degree of correlation between said paragraph and the text based on at least one of the artificial attribute, the discourse structure attribute, the paragraph length ratio, and the combination attribute;
 - important paragraph determining section operatively connected to determine an important paragraph having higher correlation with the text based on the estimated importance degree of each attribute from one or more paragraphs in the text; and
 - text output interface operatively connect to provide information of the text that is based on the determination of said important

paragraph determining section.

2. A text information generating apparatus comprising:

attribute input section operatively connected to receive at least one artificial attribute that is associated with a paragraph;

discourse structure attribute generating section operatively connected to generate a discourse structure attribute related to a discourse structure and associated with said paragraph and a paragraph length attribute related to a ratio of a number of characters of said paragraph to a number of characters of a matching pattern matched to said paragraph;

word attribute generating section operatively connected to generate word attribute related to words of said paragraph;

combination attribute generating section operatively connected to generate a combination attribute based on at least two of the artificial attribute, the discourse structure attribute, the paragraph length ratio attribute, and the word attribute;

text input interface operatively connected to receive text;

importance degree estimating section operatively connected to estimate an importance degree indicating an enhancement degree of correlation between said paragraph and the text based on at least one of the artificial attribute, the discourse structure attribute, the paragraph length ratio attribute, the word attribute, and the combination attribute;

important paragraph determining section operatively connected to determine, based on the estimated importance degree of each attribute, an important paragraph having a higher correlation with the text from one or more paragraphs in the text; and

text output interface operatively connected to provide information of the text that is based on the determination of said important paragraph determining section.

3. A text information generating apparatus comprising:

attribute input section operatively connected to receive at least one artificial attribute that is associated with a paragraph;

discourse structure attribute generating section operatively connected to generate a discourse structure attribute related to a discourse structure that may be associated with said paragraph and a paragraph length ratio attribute related to a ratio of a number of characters in said paragraph to the number of characters of a matching pattern matched with said paragraph;

combination attribute generating section operatively connected to generate a combination attribute based on at least two of the artificial attribute, the discourse structure attribute, and the paragraph length ratio attribute;

text input interface operatively connected to receive text;

importance degree estimating section operatively connected to estimate an importance degree indicating an enhancement degree of correlation between said paragraph and the text based on at least one of the artificial attribute, the discourse structure attribute, the paragraph length ratio, and the combination attribute, and to determine at least one surplus attribute from at least two of the artificial attribute, the discourse structure attribute, the paragraph length ratio, and the combination attribute ;

surplus attribute deleting section operatively connected to delete

the determined surplus attribute from the attributes utilized by said importance degree estimating section;

important paragraph determining section operatively connected to determine, from one or more paragraphs, an important paragraph having higher correlation with contents of text based on the estimated importance degree of the attribute not determined to be a surplus attribute; and

text output interface operatively connected to provide information of the text based on the determination of said important paragraph determining section.

4. The text information generating apparatus according to claim 1, wherein information of text outputted from said text output interface includes an abstract sentence based on the paragraph determined as the important paragraph.

5. An example gathering apparatus comprising the text information generating apparatus according to claim 1, wherein an incident clustering section makes one set of text from a plurality of texts using text information provided by the text information generating apparatus above.

6. A question example extracting apparatus for generating frequent text comprising the incident clustering apparatus according to claim 5;

a sorting section operatively connected to sort a plurality of question based on the gathered text; and

a determining section operatively connected to estimate frequent

test based on at least some of the sorted plurality of questions.

7. A searching apparatus comprising the text information generating apparatus according to claim 1; and

a searching section operatively connected to search for predetermined contents in text based on the information of the text.

8. A text information generating method, comprising:

receiving at least one artificial attribute and is associated with a paragraph;

generating a discourse structure attribute related to a discourse structure that is associated with said paragraph and a paragraph length ratio attribute related to a ratio of a number of characters in said paragraph to the number of characters of a matching pattern matched with said paragraph;

generating a combination attribute based on at least two of the artificial attribute, the discourse structure attribute, and the paragraph length ratio attribute;

receiving text;

estimating an importance degree indicating an enhancement degree of correlation between said paragraph and the text based on at least one of the artificial attribute, the discourse structure attribute, the paragraph length ratio, and the combination attribute;

determining an important paragraph having higher correlation with the text based on the estimated importance degree of each attribute from one or more paragraphs in the text; and

providing information of the text that is based on the

determining.

9. A text information generating method comprising:

receiving at least one artificial attribute that is associated with a paragraph;

generating a discourse structure attribute related to a discourse structure and associated with said paragraph and a paragraph length attribute related to a ratio of a number of characters of said paragraph to a number of characters of a matching pattern matched to said paragraph;

generating a word attribute related to words of said paragraph;

generating a combination attribute based on at least two of the artificial attribute, the discourse structure attribute, the paragraph length ratio attribute, and the word attribute;

receiving text;

estimating an importance degree indicating an enhancement degree of correlation between said paragraph and the text based on at least one of the artificial attribute, the discourse structure attribute, the paragraph length ratio attribute, the word attribute, and the combination attribute;

determining, based on the estimated importance degree of each attribute, an important paragraph having a higher correlation with the text from one or more paragraphs in the text; and

providing information of the text that is based on the determination of said important paragraph determining section.

10. A text information generating method comprising:

receiving at least one artificial attribute that is associated with a

paragraph;

generating a discourse structure attribute related to a discourse structure that may be associated with said paragraph and a paragraph length ratio attribute related to a ratio of a number of characters in said paragraph to the number of characters of a matching pattern matched with said paragraph;

generating a combination attribute based on at least two of the artificial attribute, the discourse structure attribute, and the paragraph length ratio attribute;

receiving text;

estimating an importance degree indicating an enhancement degree of correlation between said paragraph and the text based on at least one of the artificial attribute, the discourse structure attribute, the paragraph length ratio, and the combination attribute, and to determine at least one surplus attribute from at least two of the artificial attribute, the discourse structure attribute, the paragraph length ratio, and the combination attribute;

deleting the determined surplus attribute from the attributes utilized in the estimation;

determining, from one or more paragraphs, an important paragraph having higher correlation with contents of text based on the estimated importance degree of the attribute not determined to be a surplus attribute; and

providing information of the text based on the determining of said important paragraph.